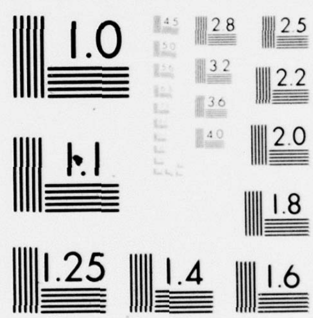
AD-A045 938 MASSACHUSETTS INST OF TECH CAMBRIDGE ELECTRONIC SYST--ETC F/G 9/4
JOINT SOURCE AND CHANNEL CODING,(U)
SEP 77 J L MASSEY N00014-75-C-1183
UNCLASSIFIED ESL-P-773 NL

1 OF 1
ADA045938

END
DATE
FILMED
12-77
DDC



MICROCOPY RESOLUTION TEST CHART

NATIONAL BUREAU OF STANDARDS-1963-A

AD A045938

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered) UNCLASSIFIED

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER

2. GOVT ACCESSION NO.

12

3. RECIPIENT'S CATALOG NUMBER

4. TITLE (and Subtitle)
Joint Source and Channel Coding

6

5. TYPE OF REPORT & PERIOD COVERED
~~Paper,~~ September 1977

14

6. ~~PERFORMING ORG.~~ REPORT NUMBER
ESL-P-773

7. AUTHOR(s)
10 James L./Massey

8. CONTRACT OR GRANT NUMBER(s)
[redacted]N00014-75-C-1183

15

9. PERFORMING ORGANIZATION NAME AND ADDRESS
Massachusetts Institute of Technology
Electronic Systems Laboratory
Cambridge, Massachusetts 02139

10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS
Program Code No. 5T10
ONR Identifying No. 049-383

11. CONTROLLING OFFICE NAME AND ADDRESS
Defense Advanced Research Projects Agency
1400 Wilson Boulevard
Arlington, Virginia 22209

12. REPORT DATE
11 Sep[redacted] [redacted]977

13. NUMBER OF PAGES
~~15~~ 12 17p.

14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office)
Office of Naval Research
Information Systems Program
Code 437
Arlington, Virginia 22217

15. SECURITY CLASS. (of this report)
UNCLASSIFIED

15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

data compression
source coding
channel coding
error correcting codes

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

The advantages and disadvantages of combining the functions of source coding ("data compression") and channel coding ("error correction") into a single coding unit are considered. Particular attention is given to linear encoders, both for sources and for channels, because their ease of implementation makes their use desirable in practice. It is shown that, without loss of optimality, a joint source/channel linear encoder may be used when the goal is the distortionless reproduction of the source at the destination. On the other hand, it is shown that in general there is an inherent and significant loss of optimality if a

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

127200

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

B

20.
source/channel linear encoder is used when the goal is relaxed to reproduction of the source within some specified non-negligible distortion.

September 1977 ESL-P-773

(Preprint of paper to appear in Communication Systems and Random Process Theory (Ed. J. K. Skwirzynski), NATA Advances Study Institutes Series, Noordhoff Int. Publ. Co., Leyden, The Netherlands. This NATO Institute took place in Darlington, England, August 8-20, 1977.)

JOINT SOURCE AND CHANNEL CODING*

James L. Massey

Visiting Professor of Electrical Engineering
and Computer Science
Massachusetts Institute of Technology
Cambridge, Massachusetts U.S.A.

ABSTRACT. The advantages and disadvantages of combining the functions of source coding ("data compression") and channel coding ("error correction") into a single coding unit are considered. Particular attention is given to linear encoders, both for sources and for channels, because their ease of implementation makes their use desirable in practice. It is shown that, without loss of optimality, a joint source/channel linear encoder may be used when the goal is the distortionless reproduction of the source at the destination. On the other hand, it is shown that in general there is an inherent and significant loss of optimality if a joint source/channel linear encoder is used when the goal is relaxed to reproduction of the source within some specified non-negligible distortion.

1. INTRODUCTION

Our aim in this tutorial paper is to treat the separability of the two basic coding functions that arise in communications, namely source coding and channel coding, first in the general case and then in the important practical case when these functions are both linear. We shall find that the desirability of joint linear source/channel coding is closely (and, to us, surprisingly) linked to the degree of fidelity specified in the reconstruction of the source at the destination.

*This research was supported by the Office of Naval Research under Contract ONR-N00014-64-C-1183.

The model of a communications system with separate source and channel coding is shown in Fig. 1.

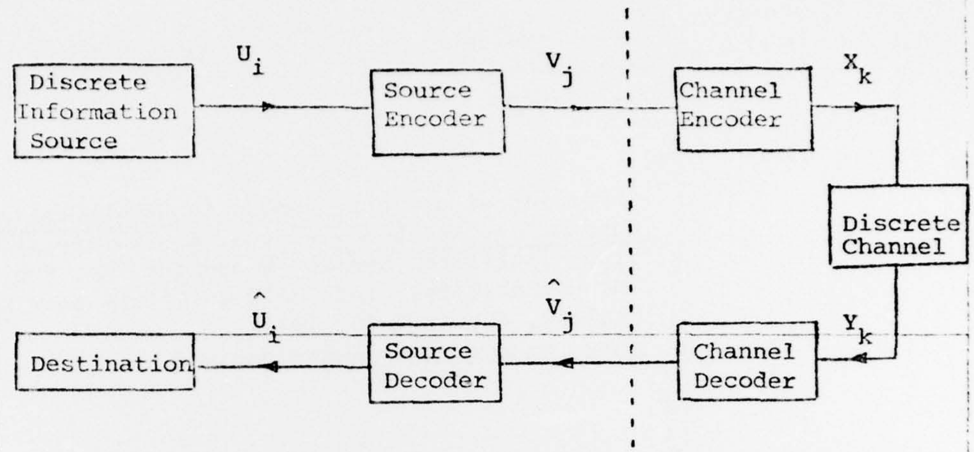


Fig. 1 A Digital Communications System with Separate Source and Channel Coding

It will be noted that there are three different subscripts on the various symbols shown in Fig. 1, namely, i, j, and k. We use this artifice to distinguish between sequences that may not be equi-numerous over a long time inverval. For instance, there may be more source output digits per second, say, than encoded source digits per second--in fact, we hope that there are many more so that the source encoder is doing well its task of "data compression". Also for instance, there may be fewer encoded source digits per second than encoded channel digits per second--we may be forced into this situation by the need to insert redundancy into the channel input digits so that the channel decoder can do well its task of "error correction".

Roughly speaking, we may use the terms "source coding", "data compression", and "redundancy removal" as synonymous. Again roughly speaking, we may use the terms "channel coding", "error correction", and "redundancy insertion" as synonymous. A wag might accuse the International Brotherhood of Information Theorists of featherbedding: it provides jobs for those who take out redundancy and jobs for those who put redundancy back in, at least when source coding and channel coding are performed separately as shown in Fig. 1. But it is a serious question to ask whether one box, a "joint source/channel encoder" as shown in Fig. 2, couldn't do a better job (or at least do the same job more economically) than does the tandem combination of the "source encoder" and "channel encoder" boxes in Fig. 1. As we shall soon be seeing, this simple question has a rather complicated answer.

In fact, one of the important results in Shannon's celebrated 1948 paper[1] was his demonstration that the source and channel coding functions are fundamentally separable in the sense that, without loss of efficiency in the use of a given channel to transmit a given source

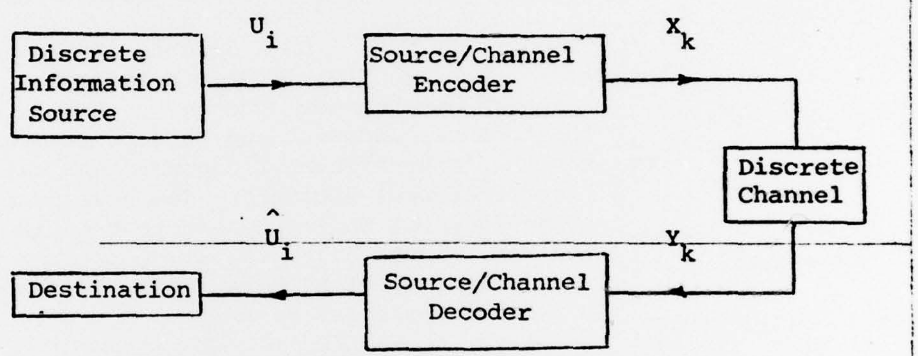


Fig. 2 A Digital Communications System with Joint Source/Channel Coding

with some specified fidelity to a destination, these two coding subsystems can be designed entirely independently. One can always design an optimum system by combining (1) a source encoder which has been designed to transform (at least, approximately) the source output into a stream of independent binary digits, each equally likely to be a 0 or a 1, and (2) a channel encoder which has been designed quite independently of the actual statistics for its input binary digits (i.e., has been designed for use with a maximum-likelihood decoder). Fano[2] has aptly commented on the significance of this fundamental separability: it means that those parts of the communications system to the right of the dashed line in Fig. 1 can always be designed, with no loss of optimality, as a system to transmit binary digits reliably. Binary digits are a kind of standard interface between the source coding world and the channel coding world, and one pays no surtax in efficiency for crossing at this interface.

As characteristic as the generality of the above-stated separability result of Shannon is the fact that his 1948 paper gives little clue as to how complex an efficient communications system becomes when the source and channel coding functions are separated as in Fig. 1. With tongue-in-cheek, we now assert:

Theorem 1: For a given efficiency (measured in number of source letters transmitted per use of the channel and fidelity (measured in the quality of the source reproduction at the destination) achievable by separate source and channel coding for a given source and a given channel, there always exists a joint source/channel coding scheme for the same source and channel that is at least as efficient, that gives at least as much fidelity, and is no more complex than the separate coding system.

Proof: Let Fig. 1 be a diagram of the hypothesized separate system. Then, in Fig. 1, draw a large box to enclose the "source encoder" and "channel encoder". Draw a second such box to enclose the "channel decoder" and "source decoder". Call the first new box the "source/channel encoder" and call the second new box the "source/channel decoder". You have just constructed a joint source/channel coding system that satisfies the assertion in the theorem. (Naturally, you might be able to build a simpler joint system that works at least as well; in fact, you might be able to build a far simpler system!)

Its triviality not withstanding, Theorem 1 does illuminate the chief attractive feature of joint source/channel coding, namely, the possible reduction in complexity compared to a similarly-performing system with separate source and channel coding. We will pursue this point further, but not without first giving a caveat: the reduction in complexity is purchased by a loss in flexibility! If one opts for a jointly coded system, he can no longer easily adapt his system later to a different source; in the separately designed system, one could continue to use the same channel coding subsystem, changing only the source encoder to the source encoder matched to the new source. Telephone companies worldwide are beginning to experience how painful this loss of flexibility can be. Most telephone systems were originally designed as a joint source/channel coding system (even if the designers were unawares that they were doing "coding") for transmitting the voice source over a narrowband channel. As more and more of their customers are changing from voice sources to data sources, the telephone companies are madly scrambling to adapt their communications brontosaurus to its new environment.

2. DEFINITIONS AND PRELIMINARIES

So that we can begin to speak more precisely as engineers should, we state here a few definitions.

A binary memoryless source (BMS) with parameter q is a device whose output is a sequence $U_1, U_2, U_3, \ldots$ of statistically independent, binary-valued random variables such that

$$P(U_i = 1) = 1 - P(U_i = 0) = q, \quad \text{all } i.$$

This is the only source that we shall consider hereafter; it is general enough for all our purposes even if it is a realistic model of only few actual information sources. When $q = 1/2$, the BMS is called the <u>binary symmetric source</u> (BSS); this very special type of BMS will play a key role in what follows. In fact, the goal of the source/encoder in Fig. 1 is to make its output a good approximation to the output of a BSS.

A <u>binary symmetric channel</u> (BSC) with cross-over probability p is memoryless channel which accepts binary digits at its input and emits binary digits at its output according to the following conditional probabilities:

$$P(Y = 1 \mid X = 0) = P(Y = 0 \mid X = 1) = p$$

$$P(Y = 1 \mid X = 1) = P(Y = 0 \mid X = 0) = 1 - p.$$

Again, although the BSC is a realistic model for only a few actual discrete channels, it is general enough for our purposes.

Next, we recall some well-known results from information theory[1,2,3,4].

Let $h(x) = - x \log_2 x - (1 - x) \log_2 (1 - x)$ (where $0 \le x \le 1$) be the usual binary entropy function. Then the <u>entropy</u> (or "rate") of the BMS is given by

$$H(U) = h(q) \quad \text{bits/letter}$$

where "letter" means a binary digit emitted by the source. According to Shannon's Noiseless Coding Theorem, $H(U)$ is the lower limit of rate, measured in encoded binary digits per source letter, for a source encoder such that the source output sequence can be reconstructed from the encoder output with an arbitrarily-small specified per-digit error probability. Equivalently, $1/H(U)$ is the upper limit of compression, measured in source letters per encoded binary digit, which can be achieved by coding schemes which convert the source output into a stream of binary digits from which the source output can be reconstructed with an arbitrarily-small specified per-digit error probability.

The <u>capacity</u> of the BSC is given by

$$C = 1 - h(p) \quad \text{bits/use,}$$

where a "use" means the transmission of a single binary digit through the channel. According to Shannon's Noisy Coding Theorem, C is the upper limit of the rate of binary digits from a BSS (which we can think of as being the output of the source encoder in Fig. 1) per channel use for a channel encoder such that there is a channel decoder which delivers the BSS digits with an arbitrarily-small specified per-digit error probability.

A very fundamental characterization of an information source is that given by its rate-distortion function. The <u>rate-distortion function</u> of the BMS is given by

$$R(D) = \begin{cases} h(q) - h(D) \quad \text{bits/letter}, & 0 \le D \le \min(q, 1-q) \\ 0, & D > \min(q, 1-q) \end{cases}$$

where D is the <u>Hamming distortion</u> defined by

$$D = \lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} P(\hat{U}_i \neq U_i),$$

i.e., D is the per-digit error probability in the source reconstruction. According to Shannon's Theorem for Coding Relative to a Fidelity Criterion, R(D) is the lower limit of rate, measured in binary digits per source letter, for a source encoder such that the source output sequence can be reconstructed from the encoder output with a distortion of D or less.

3. LINEAR CODING

We now consider the special case of linear coding, both linear source coding and linear channel coding. We begin with the latter because the relevant theory[5] is more widely known.

A [block] linear (N, K) binary channel encoder is specified by a K x N binary matrix G, of rank K, in the manner that

$$\underline{X} = \underline{V}\, G \qquad (1)$$

where $\underline{V} = [V_1, V_2, \ldots V_K]$ is the information (row) vector, and $\underline{X} = [X_1, X_2, \ldots X_N]$ is the codeword. The operations in (1), and hereafter for all matrices and vectors, are in the finite field GF(2), i.e., in modulo-two arithmetic. The code <u>rate</u> is $R = K/N$ bits/use.

It is well-known[2,3] that linear channel coding is sufficiently general to attain the performance promised by the Noisy Coding Theorem (although we hasten to add that it is only the encoder which is linear; a good channel decoder is always nonlinear!). That is, for a given $\varepsilon > 0$ and a given R such that $R < C$, there exists, for sufficiently large N, linear (N, K) encoders and appropriate decoders such that

$$\frac{1}{N} P(\hat{\underline{X}} \neq \underline{X}) \leq \varepsilon$$

when this channel coding system is used on a BSC of capacity C, regardless of the source statistics. In fact, it is known that no other type of coding can give a significantly smaller decoding error probability. Add to this the simplicity with which a linear encoder can be implemented and you will see why no one seriously proposes the use of other than linear channel encoders.

For the given G, one can always find an (N-K) x N matrix H, of rank N-K, such that

$$G\, H^T = 0 \qquad (2)$$

where the superscript T denotes "transpose". Moreover, a given vector $\underline{X}$ is a codeword if and only if

$$\underline{X}\, H^T = \underline{0}.$$

If one writes the vector $\underline{Y} = [Y_1, Y_2, \ldots Y_N]$ received over the BSC as $\underline{Y} = \underline{X} + \underline{E}$, where $\underline{E} = [E_1, E_2, \ldots E_N]$ is the <u>error pattern</u>, then it follows from (2) that

$$\underline{S} \overset{\Delta}{=} \underline{Y}\, H^T = \underline{E}\, H^T. \tag{3}$$

The (row) vector $\underline{S} = [S_1, S_2, \ldots S_{N-K}]$ is consequently called the <u>syndrome</u> because it depends only on the errror pattern $\underline{E}$ that has infected the codeword in its passage through the BSC.

It is a well-known fact in coding theory that, without loss of optimality, the decoder for a linear code can always be built in the manner shown in Fig. 3 such that the decoder first forms the syndrome and then estimates the error pattern solely from this syndrome. One should not be misled by Fig. 3; the leftmost and

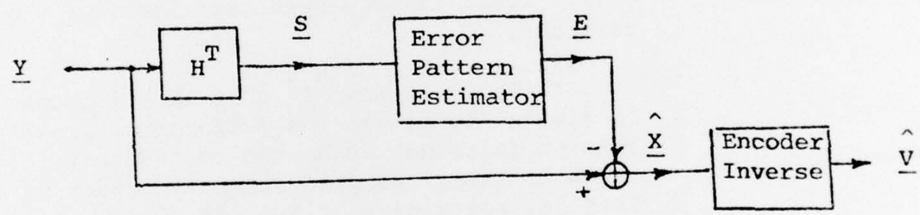


Fig. 3 A Syndrome Decoder for a Linear Code

rightmost boxes therein are linear devices and easy to implement, but the box labelled "error pattern estimator" may be unimaginably difficult to implement for very long and powerful codes.

We now turn to the description of linear source coding. A [block] linear (N, K) source encoder is specified by an (N-K) x N binary matrix H, of rank N-K, in the manner that

$$\underline{V} = \underline{U}\, H^T \tag{4}$$

where $\underline{U} = [U_1, U_2, \ldots U_N]$ is the source message, and $\underline{V} = [V_1, V_2, \ldots V_{N-K}]$ is the encoded version of the source message. (We shall place the subscript c or s on K, N, H and G whenever the context does not make it clear whether we are specifying the channel encoder or the source encoder, respectively.) Thus, the <u>compression ratio</u> of a linear (N, K) source encoder is

$$\beta \overset{\Delta}{=} N/(N-K).$$

The <u>rate</u> of this linear source coding scheme is

$$R_L \overset{\Delta}{=} 1/\beta = 1 - K/N.$$

The reason for our choosing the above notation for linear source encoding is the interpretation that we now wish to make. We first make the key observation that the error pattern $\underline{E}$ of the BSC is statistically identical to the output vector $\underline{U}$ of a BMS with parameter q equal to p. Thus, we are always free to consider that a linear source encoder treats the output of the BMS as an "error pattern" and forms the "syndrome" of this error pattern, according to (4), which syndrome is then the encoded version of the source message. Hence, we can always consider linear source coding conceptually as shown in Fig. 4 where the source decoder is an "error pattern estimator". This interpretation of linear source coding appeared first in the literature in the work of Ohnsorge[6] and has been rather fully developed by Ancheta[7].

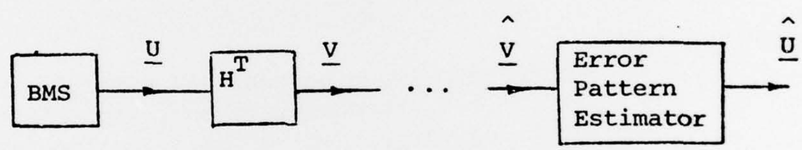


Fig. 4 The Syndrome-Source-Coding Interpretation of Linear Source Coding

4. JOINT LINEAR SOURCE/CHANNEL CODING--THE DISTORTIONLESS CASE

We now consider linear source encoding when the goal is reproduction of the source with a negligibly small (but non-zero) probability ε of digit error, so-called "distortionless coding".

Consider a BMS with parameter q where, for convenience with no real loss of generality, we take $0 \leq q \leq 1/2$. For the BSC with crossover probability p equal to q, we know there is a linear channel coding scheme (G_c, H_c) such that, for any given $\delta > 0$, it has

$$R \geq C - \delta = 1 - h(q) - \delta$$

and achieves per-digit error probability ε or less in the estimated codeword $\hat{\underline{X}} = \hat{\underline{U}}\, G_c$. For this channel coding scheme, the per-digit error probability in the vector $\hat{\underline{E}}$ of Fig. 3 coincides with that in the vector $\hat{\underline{X}}$. Thus, if we use these same two matrices as the G_s and H_s of the scurce coding scheme of Fig. 4, it follows that the per-digit error probability of the reconstruction $\hat{\underline{U}}$ is again the same, i.e., is or less. (Here we assume that the source coding scheme uses the same error pattern estimator as did the channel coding scheme.) The compression ratio achieved is

$$\beta = \frac{N}{N-K} = \frac{1}{1-R} \geq \frac{1}{h(q)+\delta} = \frac{1}{H(U)+\delta}$$

which is arbitrarily close to the upper limit of achievable compression ratios, $1/H(U)$, established by the Noiseless Coding Theorem. Thus, as has been observed by Hellman[8] and Ancheta[7], linear source encoding entails no loss of optimality when the goal is distortionless reproduction of the source.

But we now recall that linear channel coding never entails a loss of optimality. Moreover, if we have

$$N_s - K_s = K_c$$

(which can always be achieved simply by redefining the block lengths, if necessary, to be integer multiples of the original block lengths), then we can write for the tandem combination of the two linear systems

$$\underline{X} = \underline{V}\, G_c = \underline{U}\, H_s^T\, G_c.$$

It follows then that we can consider $A \triangleq H_s^T G_c$ to be the defining matrix of a linear joint source/channel encoder which operates as

$$\underline{X} = \underline{U}\, A.$$

It follows, as first observed by Hellman[8], that joint linear source/channel encoding entails no loss of optimality when the goal is distortionless reproduction of the source. Moreover, the implementation of the matrix $A = H_s^T G_c$ cannot avoid being far simpler in general that the separate implementation of the matrices H_s^T and G_c.

Example: Suppose that we are to transmit, with negligibly small distortion, a BMS with $q = .10$ through a BSC with $p = .10$. Since $h(.10) = 0.47$, it follows that a compression ratio of $1/h(.10) = 2.13$ can be approached, and that a channel coding rate of $C = 1 - h(.10) = .53$ can be approached. Thus, an overall efficiency of $(2.13)\times(0.53) = 1.13$ source letters per channel use can be approached arbitrarily closely with joint source/channel linear coding, and no larger overall efficiency can be obtained by any distortionless coding scheme. In particular, for suitably large K, we can find an $R = 1/2$ linear channel encoder specified by

$$G_c = [I_K\ P]$$

(where P is some K x K binary matrix) and a $\beta = 2$ linear source encoder

$$H_s = [P^T \; I_K]$$

such that the overall distortion is smaller than the specified small amount. But then

$$A = H_s^T G_c = \begin{bmatrix} P & P^2 \\ I & P \end{bmatrix}$$

describes a linear joint source/channel encoder which has overall efficiency $\beta R = 1$, quite close to the theoretical limit. Moreover, we see that A can be implemented quite straightforwardly from a device which implements only P, whereas implementation of G_c and H_s would each require implementation of P in separate source and channel coding. It is interesting to note that A is an N x N matrix, but that its rank is only N/2; this lack of full rank appears to be fundamental for useful linear joint source/channel encoders.

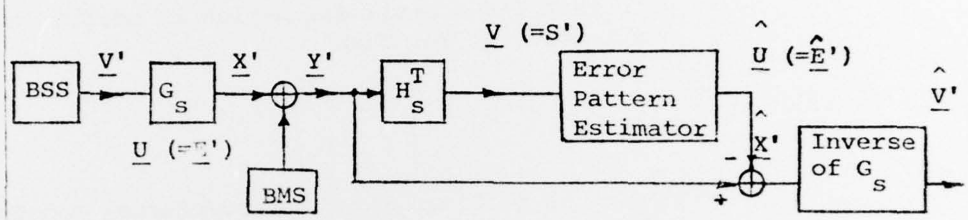
We conclude that joint linear source/channel coding is a highly attractive approach when the goal is the distortionless reproduction of the source.

5. JOINT LINEAR SOURCE/CHANNEL CODING--THE NON-NEGLIGIBLE DISTORTION CASE

With many actual data sources (e.g., with facsimile), one is often content to accept non-negligible distortion D in the source reproduction (e.g., D = 1/10). The rate-distortion function of the source specifies how such a relaxed demand on the fidelity of reconstruction can be translated into more efficient use of the channel, i.e., fewer uses of the channel for each source letter.

Following recent work by Ancheta[9], we now show that, for a given D (non-negligibly) greater than zero, the performance of linear source coding is bounded in general strictly below the compression ratio 1/R(D) which Shannon has shown can be approached arbitrarily closely by some sort of source coding.

The key (and clever) idea in Ancheta's proof that linear source encoding for non-negligible distortion in inherently suboptimal was his exploitation of the fact that a linear source encoder "cannot see" a vector which lies in the null space of the matrix H_s^T, i.e., its output is zero for any vector which could be the output of the linear device which implements the matrix G_s. Consider then the situation shown in Fig. 5, where we have merely supplemented the source coding system of Fig. 4 by adding some devices that have no effect on the latter's operation. If D is the per-digit



error probability in $\hat{\underline{U}}$ for the linear source coding scheme, we see from Fig. 5 that it is also the per-digit error probability in $\hat{\underline{X}}'$. Now, as is well-known in coding theory, given H_s, one can always choose G_s such that G_s has an identity matrix in some K of its columns. But then $\hat{\underline{V}}'$ is just the vector composed of the K digits in these K positions of $\hat{\underline{X}}'$. It follows that the per-digit error probability in $\hat{\underline{V}}'$ is at most (N/K)D. But, since this is also the fidelity with which the BSS (not the BMS!) in Figure 5 is being transmitted through the BSC created by considering the output of the BMS to be an error pattern $\underline{E}$, and since K digits of the BSS are being transmitted with N uses of this BSC with capacity $C = 1 - h(q)$, it follows from the properties of the rate-distortion function of the BSS that

$$\frac{N[1 - h(q)]}{K} \geq R_{BSS} \left(\frac{N}{K} D\right) = 1 - h\left(\frac{N}{K} D\right)$$

or, equivalently,

$$h\left(\frac{N}{K} D\right) \geq 1 - \frac{N}{K} [1 - h(q)]. \qquad (5)$$

We can put (5) into more revealing form in terms of

$$R_L = \frac{1}{\beta} = 1 - \frac{K}{N}.$$

Then (5) becomes

$$D \geq (1 - R_L)\, h^{-1} \left[\frac{h(q) - R_L}{1 - R_L}\right] \qquad (6)$$

where $h^{-1}(\cdot)$ is the inverse (made unique by restricting its values to be between 0 and 1/2) of the binary entropy function.

The significance of (6) can perhaps be most easily seen by its specialization to the BSS, i.e., to $q = 1/2$. Then $h(q) = 1$ and (6) simplifies to

$$D \geq (1 - R_L)/2. \tag{7}$$

In Fig. 6a, we have plotted both the bound (7) on the attainable distortion D of a linear source coding scheme of rate R_L for the BSS, together with the rate-distortion function $R(D) = 1 - h(D)$ of the BSS. This figure clearly illustrates how far away from optimal a linear source coding system must be when a non-negligible D is specified. For example, with $D = .11$, $R(D) = .50$ but $R_L = .78$. Thus the linear scheme can have at best $\beta = 1/R_L = 1.28$, compared to the compression ratio $1/R(D) = 2$ that can be approached by more general source coding schemes.

A similar interpretation can be made from Fig. 6b where we have shown the rate-distortion function R(D) for the general BMS and also the corresponding bound on R_L from (6).

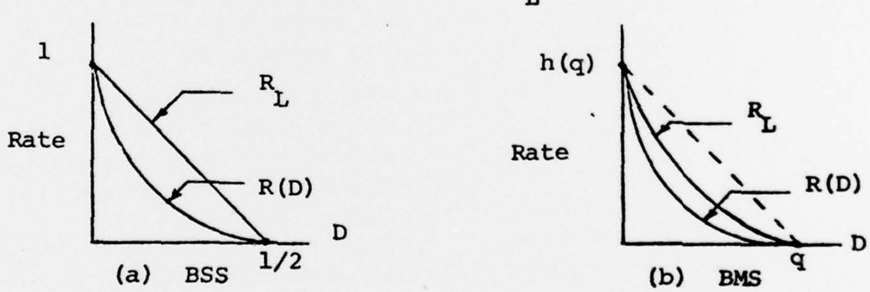


Fig. 6 Bounds on the Achievable Rate R_L with Linear Source Coding

Ancheta[9] actually has a lot more to say about the non-optimality of linear source coding with non-negligible distortion, but we shall leave the rest for him to tell in his own publications, except to mention his conjecture that the achievable R_L is actually more strictly bounded away from R(D) according to the dashed line shown in Fig. 6b.

We now give a simple argument to show that the inherent lack of optimality of linear source coding in the non-negligible distortion case implies in general an inherent lack of optimality for linear joint source/channel encoding in the non-negligible distortion case.

Suppose that the $N_s \times N_c$ matrix A describes a linear joint source/channel encoder, for a BMS and BSC, which achieves distortion D (where D is not negligibly small). Suppose that A has rank r. Then one can always find an $r \times N_s$ matrix H_s of rank r and $r \times N_c$ matrix G_c of rank r such that $A = H_s^T G_c$. Thus, we can consider the matrix H_s^T as describing a linear source encoder and the matrix G_c as describing a linear channel encoder; the original linear joint source/channel encoder is equivalent to separate encoding with these derived linear encoders.

Let D' be the best obtainable distortion when the BMS is reconstructed directly from the output of the linear source encoder H_s^T. It follows that $D' \geq D$, because the best service which the channel encoder G_c can provide is to permit perfect transmission of the source encoder output to the best source reconstructor. Hence, the rate R_L of the linear source encoder must satisfy (6) for the given distortion D.

The overall efficiency of the linear separate source/channel coding system (and hence also of the entirely equivalent original linear joint coding system) is $\beta R_c = R_c/R_L \leq 1/R_L$ source letters per channel use, where the inequality follows from the fact that $R_c \leq 1$. On the other hand, there exist coding systems whose overall efficiency approaches C/R(D) source letters per channel use, where C is the capacity of the BSC and R(D) is the rate-distortion function of the BMS. Thus, when, for a given D, the bound (6) specifies an R_L such that $R_L > R(D)/C$, then there is an inherent loss of optimality when linear joint source/channel encoding is used. In other words, when the bound (6) gives an R_L which exceeds R(D) by a factor of more than 1/C, then linear joint source/channel encoding is sub-optimum.

Example: Consider the BSS together with the BSC having p = .10, and suppose that D = 1/4 is specified. Then, R(D) = h(1/2)-h(1/4)= .19. From (6), we find $R_L = .50$. Thus, R_L is (.50)/(.19) = 2.63 times as great as R(D). But 1/C = 1.89. Because 2.63 > 1.89, it follows that a linear joint source/channel coding system must be sub-optimum. To put it another way, any such linear joint coding system has an efficiency of at most $1/R_L = 2$, whereas there exist more general coding systems whose efficiency approaches C/R(D) = 2.79 source letters per channel use.

We should point out in closing that a joint linear source/channel coding system can sometimes "accidently" be optimal when R_L, as given by (6), exceeds R(D) by a factor of only 1/C or less.

In the above example, if we had taken $D = .10$ rather than $D = 1/4$, we would have found $R_L = .80$ and $R(D) = .53$ so that $R_L/R(D) = 1.51 < 1/C = 1.89$. $C/R(D)=1$ is the maximum approachable efficency. But the "straight wire" encoder, which merely transmits the BSS output directly over the channel, has efficiency 1 and distortion $D = .10$. We can consider this trivial but optimum coding scheme as the linear joint source/channel coding scheme with $A = 1$. [The reason for this accidental optimality is that the given BSC happens to be the appropriate "forward channel" for the given distortion D and the BSS, cf. Berger[4]]

References

1. Shannon, C.E., A Mathematical Theory of Communication, Bell System Technical Journal, 27, 379, 1948.

2. Fano, R.M., Transmission of Information, M.I.T. Press, Cambridge, Massachusetts, 1961. (page 3)

3. Gallager, R.G., Information Theory and Reliable Communication, John Wiley and Sons, New York, 1968.

4. Berger, T., Rate Distortion Theory: A Mathematical Basis for Data Compression, Prentice-Hall, Englewood Cliffs, New Jersey, 1971.

5. Peterson, W.W., Error-Correcting Codes, M.I.T. Press, Cambridge, Massachusetts, 1961.

6. Ohnsorge, H., Data Compression System for the Transmission of Digitalized Signals, Proceedings IEEE International Conference on Communications, 485, 1973.

7. Ancheta, T.C., Jr., Syndrome-Source-Coding and its Universal Generalization, IEEE Transactions on Information Theory, 22, 432, 1976.

8. Hellman, M.E., Convolutional Source Encoding, IEEE Transactions on Information Theory, 21, 651, 1975.

9. Ancheta, T.C., Jr., Bounds and Techniques for Linear Source Coding, Ph.D. Thesis, Department of Electrical Engineering, University of Notre Dame, Notre Dame, Indiana, August, 1977.